REENGINEERING OF THE CANADIAN MONTHLY RESTAURANTS, CATERERS AND TAVERNS SURVEY USING STATISTICS CANADA GENERALIZED SYSTEMS

Charles Tardif and Ron Carpenter,
Statistics Canada, Canada

ABSTRACT

The Monthly Restaurants, Caterers and Taverns Survey (MRCTS) has been running in its current format since 1980. The objectives of MRCTS are to provide current information on sales, trends and changes in the food- and beverage-serving industry. Up until May 1995, all production was run using an in-house system developed especially for MRCTS. This system met all the needs of the survey at the time of its development. However, due to new requirements and a few problems encountered during the regular production, the time was right to redesign the survey systems in order to improve its effectiveness and its reliability. All aspects of the survey have been redesigned. Generalized Systems developed at Statistics Canada have been incorporated in the regular production. The redesign will greatly facilitate and speed up the entire maintenance of MRCTS and will reduce the cost inherent to the survey, since the systems development cost are reduced to a minimum. This will result in better estimates produced at a lower cost.

KEYWORDS

Restaurants, Reengineering, Generalized Systems, Survey Improvements.

INTRODUCTION

In 1996, the Annual Research Conference (ARC) focuses on methodologies and technologies for improving surveys and censuses. This paper, written for a presentation at this conference, describes the reengineering of the Monthly Restaurants, Caterers and Taverns Survey (MRCTS). Early results obtained with this redesign are also discussed.

The objectives of MRCTS are to provide current information on sales, number of locations, trends and changes in the food- and beverage-serving industry. Information from this survey is used by members of the industry to compare their individual growth with that of the province and to create marketing techniques. It is also used by Federal and Provincial governments to develop fiscal policies and to measure and forecast growth in the industry. Other major users include the Restaurant Association of Canada, universities and tourism associations and agencies that are involved in the development of the food and beverage industry.

The sampling units in the frame are businesses that are primarily concerned with the serving of food and beverages. These units are classified into five categories, or Kind of Business (KOB):

1) licensed restaurants,

2) unlicensed restaurants including drive-ins,
3) take-out food shops including refreshment stands,
4) caterers, and
5) taverns, bars and night clubs.

Until May 1995, all production was run using systems developed exclusively for MRCTS in the early 1980's. Since then, the technologies to handle the different aspects of a survey have greatly improved. With new survey requirements and a deterioration of the frame, a review of the survey was timely. Originally, only the collection procedures, the sampling procedures and the production of the estimates were to be changed. Later, it was realized that updating all components of MRCTS would improve the survey significantly.

As mentioned earlier, the redesign of MRCTS became necessary mainly due to the availability of the new technology and the deterioration of the actual systems. In order to ensure that this redesign will be efficient, the main objectives have been identified as:

1) To improve and maintain the quality of the data,
2) To modernize the systems to facilitate their use,
3) To decrease the time required to produce the estimates,
4) To reduce manual procedures wherever possible, and
5) To save resources and money.

The methodology and technologies used for the redesign are discussed for each of the following components of MRCTS:

1) Frame maintenance,
2) Sampling process,
3) Data collection,
4) Edits and imputation,
5) Estimation of totals and associated variances, and
6) Review and analysis

FRAME MAINTENANCE

The Canadian Business Register (BR), developed and maintained at Statistics Canada, is used to delineate the frame for the MRCTS population. The BR is a central repository of information on all employer businesses in Canada and exists to supply lists for all business surveys in Statistics Canada. Each business in the BR is coded using the 1980 Standard Industrial Classification (SIC) to identify its major activity. For MRCTS, businesses coded to 9211, 9212, 9213, 9214 and 9221 are relevant and in-scope to the survey. These SIC codes correspond to the five KOB categories described earlier.

Each month, a Statistical Universe File (SUF) is created by extracting all businesses in-scope for MRCTS from the BR. There are currently about 60,000 in-scope units for MRCTS. The SUF is used by the Generalized Sampling System (GSAM) to maintain the sampling frame. The SUF is compared with the previous month’s SUF to identify births and deaths in the population; new units are stratified and dead units are removed. The frame maintenance is used to prepare information for the

sample selection step.

One of the on-going tasks for this survey is to complete a fully integrated hookup with the BR; this would allow MRCTS to provide automated survey feedback to the BR. Using the BR as a frame source will allow the survey to feedback frame changes and to take advantage of frame updates provided by other hooked-up surveys without having to maintain its own divisional frame.

SAMPLING PROCESS

The Generalized Sampling System (GSAM) is also used each month to select and maintain the sample. The current resources allow the survey to maintain a sample of about 3,000 units each month. The selection process uses a stratified simple random sampling design with a fixed sampling fraction (f_h) in each stratum. The population is stratified by province, KOB and size; this results in 120 strata (12x5x2). The size measure is based on annual revenue and is used to determine the split between the large and small units. The large units are sampled with certainty (take-all strata, i.e. $f_h = 1$); the small units fall into the take-some strata ($0 \leq f_h \leq 1$).

The GSAM methodology has been adapted from collocated sampling procedures. It has been modified to handle frame changes, population births and deaths and sample maintenance through successive sampling periods. GSAM handles the MRCTS survey requirements and it will also handle sample rotation when it will be introduced to the survey.

The sample selection is as follows. Initially, all units in a stratum are assigned equally spaced random numbers between [0,1]. Units whose random number fall in the interval [0, f_h] are selected in the sample. New units are independently assigned random numbers using the same method. This results in a representative sample of both new and existing units.

Although the overall sample size is not fixed, it does not vary too greatly from month to month; about 30 to 40 new units are sampled each period across the 120 strata. The new sample units are often offset by a corresponding number of deaths. If ever the sample grows too large, GSAM can be used to decrease its size while maximizing the retention of the existing sampled units.

DATA COLLECTION AND ON-LINE EDITS

The data for MRCTS is collected mainly by telephone, and also by mail or fax in the Regional Offices (RO) of Statistics Canada. Before the redesign, preprinted data sheets were used to record the information provided by participating firms. Under the redesign, the Data Capture and Collection (DC2) system which was developed at Statistics Canada is used to collect the data. The main purpose of this system is to capture and validate the data collected by the interviewers.

Each month, interviewers in the RO are given a list of businesses to contact. Each business is identified with a unique statistical company number. Before the first contact, the following information is already captured in DC2 for each business:

Reference period,
Statistical Number,

Take-all/take-some identifier, and
Birth identifier.

The first time a business is interviewed, the following information is asked and captured:

Legal name,
Operating name,
Address,
Contact name, and
Reporting arrangements (best time to call, language of preference, etc.).

On subsequent contacts, this information is displayed to the interviewer and is updated as required.

The following information is captured each month:

First and last date of the reporting period,
Province and KOBs in which the business operates, and
Sales and number of locations for each combination of province and KOBs in which the business operates

During the data capture, any updates are made to the data reported in the previous period. After the reported sales and number of locations are entered, a status code is derived. This code indicates whether or not the business is operating, or whether or not it passes the edits which are described below. Finally, there is a field to capture any comments that might be useful for processing the data.

Each record that is captured by DC2 goes through a series of edits. Some edits ensure that data such as reporting period, previous reporting period, province, KOB and status code are properly entered. Validation of the status code with the presence or absence of data in the sales and number of locations fields is performed. Consistency edits, which are described in detail below, are applied on the sales and the number of locations data, if the sales are reported on a monthly basis.

For units previously in sample, the consistency edits are based on a historical trend. This method uses the values reported in the previous year and the values reported in the previous month. Two measures of variation are calculated; one utilizes the values from previous year versus the reference month and the other one uses the values from the previous month versus the reference month. A unit passes the edits if at least one of these measures falls within the boundaries of acceptance. These edits have been designed in such a manner as to take into account the seasonality of the restaurant industry and the major changes that might occur within a year. For example, consider a business that always has a large increase in one specific month during the year. It would be incorrect to conclude that a unit fails the edits on the basis of the monthly trend if this unit has a similar revenue pattern every year. It would also be incorrect to say that a unit fails the edits based solely on the yearly trend if it is known that a major change (large expansion or closure of many locations) occurred within the last twelve months. However, if one of the historical values (previous year or previous month) is missing, the test is based on the one

available measure of variation.

Let $r_i = \frac{y_i}{y_{i-1}}$ be the ratio of the current period to the historical period

where y_i = (sales for the current period) / (100000 * number for locations of the current period) and y_{i-1} = (sales for the historical period) / (100000 * number of locations for the historical period).

Let

$$s_i = \begin{cases} 1 - \frac{r_m}{r_i} & if\ 0 < r_i < r_m \\ \frac{r_i}{r_m} - 1 & if\ r_i \geq r_m \end{cases}$$

where r_m represents the median ratio.

Then, the period to period measure of variation is given by:

$e_i = s_i * (max_i)^u$ where $max_i = \max[y_i , y_{i-1}]$ and u is the curvature parameter.

The parameters to construct the intervals of tolerance were derived using MRCTS data from December 1992 to December 1993. Due to seasonality of the restaurant industry, the monthly median ratio was set to 1.0. The curvature parameter, u, was set to 0.5 in order to accept greater relative variations for small units. The y_i and y_{i-1} in the formula represent standardized sales. Two reasons justify the division of the sales values by the number of locations. Firstly, we decrease the importance of a large variation in sales caused by a variation in the number of locations. The second, and more important reason, is that large and small establishments can be compared. Indeed, if no adjustments were made for the number of locations, a record with a high number of locations would be more likely to fail the edit than one with only one location, since its sales are greater and the importance of the factor $\mathbf{max_i}$ would be too predominant.

For new sample units, only take-some businesses are edited; to pass the edits, these units must have sales less than or equal to a maximum province by KOB value. The maximum values were calculated using MRCTS historical data.

These on-line edits give the interviewer the opportunity to immediately correct any errors that might be present, whether it be a capture error or an incorrect response. Follow-up is thus eliminated and the response burden is consequently reduced. If the respondent confirms that his answer is accurate but it fails the edits, a flag is used to identify the reason for the edits failure. After the data has been collected in the regional office, it is then transmitted to the Head Office.

EDITS AND IMPUTATION AT THE HEAD OFFICE

Once all the data is received at Head Office, edits and imputation are required to obtain a

complete file for the production of the estimates. First, non-monthly sales figures are converted to a monthly basis. This conversion is function of the number of days in the reporting period, the number of days in the reference month and the reported sales figures. After all the data is converted, the entire file is edited; the methodology used for editing the data is the same as the one implemented in DC2. Processing the edits again is necessary as the non-monthly reporters were not edited. Each record is then classified into one of the following three categories:

1) records with missing values which require imputation.

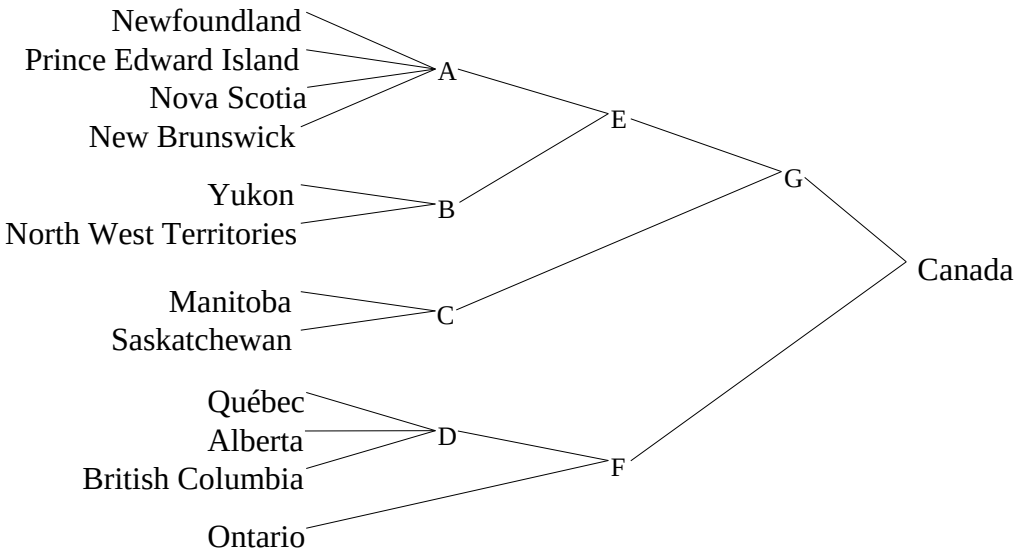
2) records with values which did not pass the consistency edits. Although these records might have valid values, their trend is too different from the average, to be used in computing means for imputation purposes. These units, however, retain their values.

3) records with values which pass the edits. These records will be used to compute means for imputation purposes.

After the edits have been completed, missing values for sales and the number of locations are imputed. The imputation strategy for these two variables is given below.

Sales:

Five groups of imputation are identified to each sample units. These groups are created to ensure that units included in the same group have similar characteristics. A group consists of a combination of province, KOB, historical status and type. The historical status is identified as yearly if yearly historical data is available, as monthly if yearly data is not available but the unit was in sample the previous month and as a birth if the unit is new to the survey. The type is either take-all or take-some. A minimum of 10 records must be present in a given group to perform the imputation; otherwise the units are combined with the units comprising the next higher level. The collapsing will continue until either the number of units is sufficient or the top level is reached. The top level corresponds to all units within the KOB, historical status and type across all provinces. The diagram below shows the collapsing pattern used by MRCTS.



At the first level of the diagram, before any collapsing, there are 360 groups (12 regions by 5 KOB's by 3 historical status by 2 types). After all the collapsing has been performed, there are 600 different potential groups of imputation for MRCTS. It is expected that most of these groups of imputation will not actually be used for two main reasons. First, with the target response rate of 95%, it is expected that there will not be a lot of values to impute. The second reason is that for most groups, there should be a sufficient number of units in the group at the first level and consequently, at the higher levels, all the values will already be imputed.

The imputation methods are based on historical trend or on current mean. The yearly trend is used for a unit if correct yearly historical data is available for that unit. Otherwise, the monthly trend is used if the monthly historical data is available and correct. The calculation of the imputed values using trends is as follows.

1) A unit which requires imputation, as well as the appropriate imputation method, is identified.

2) The groups in which the unit is contained are identified. For each level of collapsing, there is a group associated to that unit.

3) If the number of units in the group at the first level of collapsing is greater than or equal to 10, all units with valid data are used to compute the sales means for the current month and for the historical period. The ratio of the current mean of the group over the historical mean of the group is then multiplied by the historical value of the unit to obtain the imputed value.

4) If the number of units in the group is less than 10, continued collapsing to the next level is used until the required minimum number of units in the group is reached. Then, the imputed value is calculated using the process described in 3.

If the number of units in a group is insufficient at the highest level of collapsing (Canada level)

for the yearly trend, then the monthly trend is used instead. This is likely to occur during the first year of the redesign since many units are included in the sample for the first time.

For the newcomers to the survey, the imputation method is as follows.

1) A unit which requires imputation is identified and the current mean method is used.

2) The groups in which the unit is contained are identified. For each level of collapsing, there is a group associated to the unit. The previously described groups of imputation are regrouped based on the same province (or group of provinces), KOB and type. These groups are the same as those for the trend imputation, except that there is no distinction based on the historical status. This is required since it would be hard to obtain enough units in the groups if only the births were used, due to the small number of births each month.

3) If the number of units in the group at the first level of collapsing is greater than or equal to 10, all units with valid data are used to compute the imputed sales mean.

4) If the number of units in the group is less than 10, continued collapsing to the next level is used until the required minimum number of units in the group is reached. Then, the imputed value is calculated using the process described in 3.

It is well known that the imputation using the current mean method is not the best way to impute a value for a unit. Since there is no other information available for MRCTS, this method is the only possible choice. However, a unit for which no data was available in the previous month may be able to provide data for the following reference period. In these cases, the imputed value using the current mean method will be replaced by a new imputed value which will use the information available. This method can be called “backward imputation” since the value of a given month is used to impute a value for the preceding month. The methodology used for this type of imputation is the same as the one described for the regular historical monthly trend, except that the value of the previous period is imputed instead of the value of the current period.

Number of locations:

The imputation strategy for a missing value for the number of locations is quite simple. If the number of locations for a given unit is available from the previous month, then that value is used to impute for the current month. If this information is not available, then data from the previous year is used if it exists. If no such information is available, a value of 1 is imputed for this variable.

ESTIMATION OF TOTALS AND ASSOCIATED VARIANCES

Once the data have been edited and imputed, the sample file is ready for estimation. Estimates are produced for each combination of province by KOB. Totals by province and by KOB are produced, as well as totals at the Canada level. The following estimates are produced each month for publication:

1) Preliminary estimates of sales and number of locations for the reference month,
2) Revised estimates of sales and number of locations for the month prior to the reference month,
3) Year-to-date estimates of sales.

The Generalized Estimation System (GES), developed at Statistics Canada, is used to compute these estimates and their associated variances and coefficients of variation. The sample design for MRCTS is a stratified one-stage cluster sample with simple random selection. A cluster represents an establishment, which is defined as the smallest separate accounting entity capable of reporting all elements of basic industrial statistics (i.e., employment, salaries and wages, sales, all expenses and inventories). An establishment can report sales in more than one province or KOB. If this is the case, each combination of province/KOB is recorded separately. All combinations of province/KOB for a same establishment are the units of a cluster. If there is only one combination of province/KOB for an establishment, the cluster has only one unit.

The general linear regression model (GREG) is used by GES to produce the estimates. Three main steps are involved in the calculation of the estimates. First, the design weights must be calculated for each cluster. The design weights are the inverse of the probability of selection for the cluster and are calculated by GES using the cluster sample file and the population counts file. The population counts file contains, for each stratum, the number of clusters in the population. Then, the weight calculated for the cluster is assigned to all the units in the cluster.

Secondly, GES allows the use of auxiliary information to improve the estimates and to reduce the variance. The expansion estimator model, which is a special case of the GREG model, is used to produce estimates. For MRCTS, the only auxiliary information that is presently available is counts. The g-weights, which are calculated using auxiliary information, are thus one for every unit in sample.

Thirdly, the estimates themselves must be calculated using domain estimation. A domain is a subpopulation of elements for which estimates of specific characteristics or parameters are required. As mentioned earlier, for MRCTS, the domains of interest are combinations of province by KOB, province, KOB and Canada level. Descriptions of the calculation of monthly and year-to-date estimates are given below.

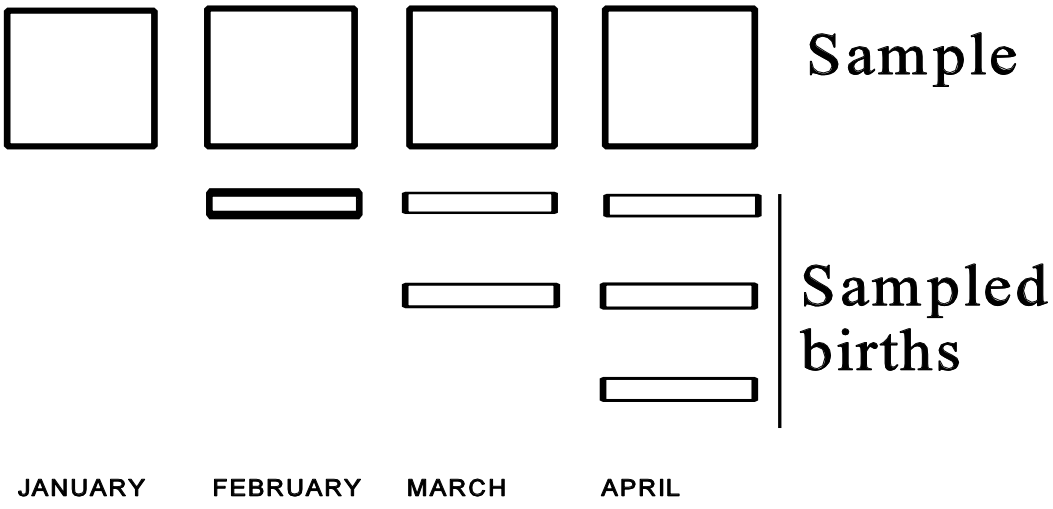
Monthly estimates of sales and number of locations:

Each month, the MRCTS sample is composed of births and previously selected units. For the purpose of calculating estimates, there is no need to know when a unit entered the sample. The parameters to estimate (sales or number of locations) can be derived simply by multiplying the monthly figures by their design weight and their g-weight. Then, adding up all the figures within

the same domain will give the required domain estimates. However, for the calculation of the variance, the situation is different. Since there are births in the universe and some of them are sampled, all units are not subject to selection at the same time. Consequently, the appropriate technique to calculate the variance, from a theoretical point of view, would be to group the units based on their time of entry into the universe. Then, the variance can be computed separately for each of these groups and added up to obtain the overall variance. In GES, each group must be processed separately; this is a major inconvenience since the number of groups increases after each new survey cycle. On a practical level, this would be difficult to implement using GES. The practical solution is to compute the variance with no regard to the time of entry into the sample. The result is an overestimation of the variance, since some covariance terms are calculated for independent units. Due to the relatively small number of births each month, calculating the variance in this way will have only a small impact on the variance, i.e. the overestimation of the variance will be very small. Along with the estimates and the variances, GES also provides coefficients of variation (CV). A CV is defined as the ratio of the square root of the variance over the estimate. Being unit-free, CVs are good measures of the variability of the data and are published with the estimates.

Year-to-date estimates of sales:

The principle of calculating the year-to-date estimates and their variance from the theoretical point of view can be understood more easily by using an example and the following diagram. For the purpose of the example, assume that a new survey starts in January of a given year and the cumulative estimates from January to April are required.



In the figure, the first row of blocks represents all units in sample from January.
The second row represents the births in February. These units then remain in sample.
The principle is the same for the next two rows. The third row represents the births that occur in March and finally, the fourth one, the births in April.

The characteristics of the diagram are:

Each row, taken separately, is an independent sample from the others; the arrival time of the units in the universe is different from one row to another, but is the same within the row.

Within a row, the units are the same from month to month and consequently, they constitute dependent samples. Thus, there is a covariance term between each pair of blocks within a row.

For each domain of interest, the theoretical calculation of the estimates and the variance is as follows:

For each row r, a new variable Z_r is created, which is the sum of sales for a unit over all months included in the row. By proceeding in this way, all the covariance terms within a row will be incorporated for the purpose of calculating the variance of the year-to-date estimates. Then, the variance of each Z_r, $Var[Z_r]$, is computed for each row r.

Finally, the year-to-date estimate, Z, is simply the sum of all Z_r within the domain of interest. Similarly, the year-to-date variance, $Var[Z]$, is the sum of all the $Var[Z_r]$, since all the rows are independent.

It is important to follow these rules for the calculation of the variance on the theoretical point of view but the estimates can be calculated all at once without altering the results.

The method described above is theoretically exact but is difficult to implement. After each cycle, the number of rows increases and it becomes harder to handle the large number of rows since calculations on each of them must be run separately. As stated earlier, a more practical solution is to calculate the estimates and the variance with no regard to the time a unit was first sampled. Using again the example with the diagram, the estimates and the variance are calculated as follows, for a domain of interest:

All rows are regrouped. From a practical point of view, this means that the entire sample file is used. Then, a new variable Z is created, which is the sum of sales for a unit over all months. All the Z are then summed to compute the year-to-date estimate for the domain of interest. The variance of the year-to-date estimate is calculated using the Z variable, for all Z in the domain of interest. Finally, along with the estimates and the variance, GES also calculates the CVs which are published with the estimates.

The impact of this method is an overestimation of the real variance. However, due to the small number of births each month, this overestimation is likely to be fairly small.

Other considerations about the production of the estimates:

Outlier detection:

One of the objectives of any survey is to provide the best possible values for the parameters to be estimated. One way to improve the quality of the estimates is to include in the survey process a feature to detect possible outliers. For MRCTS, the outlier detection is done by looking at the top contributors by domain. If a given unit accounts for a disproportionately high percentage of the

total domain estimate, then that unit has its sales value reviewed by subject-matter specialists. The value is either confirmed as accurate or corrected if there is an error. If the value is confirmed and the unit is a take-some, its weight is changed to one and the other units in its stratum are reweighted.

Effect of the imputation on the variance:

As described earlier, missing values are imputed to obtain a complete file for estimation. It is well known that treating imputed values as observed values can lead to an underestimation of the true variance. For MRCTS, the response rate is usually high and as such, not many values have to be imputed. The actual version of GES does not give the user the opportunity to treat the imputed values separately for the calculation of the variance. Treating the imputed values outside GES would complicate the processing of MRCTS, which must remain as simple as possible. For these reasons, no special treatment will be introduced for the variance calculation. However, if the response rate for MRCTS deteriorates in the future and a higher percentage of imputation is required, then procedures will have to be considered to avoid an important underestimation of the variance. If a future version of GES incorporates a feature that allows the treatment of the imputed values in the variance calculation, it will have to be examined in order to determine if it can be used for MRCTS.

Post-stratification:

Post-stratification is a common technique for improving the precision of estimators by using data items not available at the design stage of the survey. For MRCTS, more up-to-date information may be available for the population counts at the time of estimation, especially for the production of revised figures. This information can be used and incorporated in GES to improve the estimates. The g-weights will then be adjusted to reflect the new information and produce better estimates.

REVIEW AND ANALYSIS

The review and analysis of the estimates is the final step to ensure the reliability of the published results. Subject-matter specialists examine the estimates to ensure their accuracy. Historical trends, levels of sales and number of locations and important variations are the key factors that are reviewed. Any unusual situations that are found are investigated to determine the cause. If it is found that an unusual situation resulted from an error in the data, the error is immediately corrected and the estimates and the coefficients of variation reproduced.

During the review and analysis stage, we have to keep in mind that the estimates produced for MRCTS are derived from a sample survey and, as in any survey, they are subject to errors which fall into two categories: sampling errors and non-sampling errors. The quality of the estimates is influenced by the combined effect of these two sources of errors. These errors are described below, along with their impact on MRCTS.

Sampling errors

Sampling errors occur because data are collected for only a part of the population, instead of the whole population as in the case of a census. They depend on factors such as the sample size, the variability in the population, the sample design and the estimation method. For example, for a given sample size, the sampling error will depend on the choice of sampling units, the stratification method, the allocation of the sample and the selection method. For MRCTS, the estimates at the Canada level are quite good, they are less reliable at the provincial level and may be close to unreliable at the provincial/KOB level for some cases, because of the high CV's attached to some of these estimates. These last estimates must consequently be used with caution.

Non-sampling errors

These errors are present whether a sample or a complete census of the population is taken. Non-sampling errors stem from several sources, the most important of which are described below.

Coverage error. Two types of coverage errors are possible. There is under-coverage if the sampling frame does not cover the entire target population. This means that the sample cannot represent the entire population. There is an over-coverage error if a unit is represented more than once on the frame. For example, if a company on the frame is part of a large chain, this company is represented by its own record and by the chain's record.

This type of error can be quite important for MRCTS. It is well known that the restaurant business is one of the most volatile industries. Restaurants constantly go in and out of business within a short time span. Restaurants frequently acquire one another and consequently, one stops existing under its own legal name. Unfortunately, there is a delay before the frame can be updated for such changes and consequently, problems of under-coverage or over-coverage occur.

Response error. This error may be due to the inability or unwillingness of the respondent to provide correct information. It may also stem from the tendency of interviewers to explain questions or interpret responses differently.

For MRCTS, response errors can occur if the respondent does not provide the appropriate reporting period associated with its reported sales. For example, if the respondent is asked for its monthly sales for a given month but provides its annual sales without mentioning this fact, it is likely that there will be a large impact on the estimates, especially if the unit is a take-some with a large weight. Other examples of response errors are companies operating in more than one field of activity (restaurant industry and other fields) that report sales for all their activities, including activities not of interest to MRCTS. Finally, response errors also occur if the respondent is unable to classify correctly its KOB and reports in an other KOB.

Non-response. Some respondents may refuse to answer questions or be unable to respond, while others may be too late in responding. Imputation methods are used to fill the gaps left in the estimate file by these non-respondents. The extent of error due to imputation is usually unknown and is very much dependent on any characteristic differences between the respondent group and the imputed units in the survey. This error generally increases with the non-response rate and attempts are therefore made to obtain as high a response rate as possible.

The impact of non-response for MRCTS is more important for birth units, since they are imputed using the current mean method. However, since the number of sampled births for MRCTS is

relatively small, it is likely there will not be a large impact caused by the non-responding births. The other non-responding units, previously in sample, are imputed using historical trend. This is a more reliable way to impute since we use information on the specific unit to impute. However, if the number of units to impute is fairly large, there can be an impact on the estimates.

Processing error. This type of error may occur at various stages of processing such as coding, data entry, editing, imputation, estimation and tabulations.

For MRCTS, as for any other survey, processing errors are likely to happen at all stages, especially in the beginning of the redesign, when most systems are new to those running them.

Non-sampling errors are difficult to measure. Within the limitations of budget and human resources available, all attempts are made to minimize this type of error in producing the estimates for MRCTS.

THE FIRST FEW MONTHS OF OPERATION UNDER THE NEW REDESIGN

All systems implemented for the re-engineering of MRCTS are now operational and estimates have been published using them. However, like any other important projects, a few unexpected problems occurred in the early stages.

Due to a lack of time and resources, no reconciliation was done between the old and the new frame. It meant that the new frame was used “as is”. The consequences were that, for the first month of the redesign (May 1995), there were a high percentage of dead units in sample and a very high percentage of births (77%). This caused an unexpected increase in the work required to collect the data in the RO at start up. Consequently, only half of the respondents were contacted the first month. However, in the second month, most units were contacted and were asked for their data for the current and previous months. In the following months, the interviewers came up to speed in terms of collecting the data.

In the meantime, the people responsible for the development of the systems had been shifted to other urgent projects. One of the responsibilities of this team is to receive the data from the RO and to create a file that would be used for the processing at the HO. Because this task had not been completed for a few months, the data transmitted from the RO had not been reviewed. Obviously, this situation caused a delay for the entire processing at HO.

Once this situation was resolved and the data became available, the edits and imputation processing was performed. It was discovered that the status codes used to categorize the data were not always accurate. After an investigation, it was found that some edits were not applied as planned. The problems were identified, the situation had been corrected and the status codes became meaningful. The impact of this problem was that some records with miss-entered sales values were not identified during the capture process. These units were fortunately identified at Head Office and corrected with the appropriate values after follow-up.

DC2 allows the interviewer to capture comments that are reviewed by subject-matter people at Head Office. In the first few months, a very high number of comments, many being irrelevant to the survey processing, were captured in this field. This caused the processing to slow

considerably and many months of data had to be processed at the same time. As a result, specifications on how to determine which units require review have been changed. These new specifications helped reduce the number of comments to be reviewed and speed up the survey processing.

After a few months of data were processed, other problems related to the collection procedure were identified. The files received at HO had missing units, based on the expected returns. The data captured at the RO had been reviewed and the missing units were in the DC2 system. Consequently, it was determined that transmission problems had occurred during the processing. The reasons for this situation still needs to be resolved to ensure it will not happen again.

All these unexpected problems caused delays in the production and the publication of the estimates. The estimates have been reproduced several times but they have not been immediately released because of the problems encountered. The Canadian restaurant industry and some respondents have inquired about the non-availability of the estimates, at the early stages of the redesign. The subject-matter specialist explained that a major re-engineering was underway and that it had caused the delays. The interest of these people in the MRCTS estimates shows the importance of the survey, since many members of the restaurant industry use the figures to determine their market share.

CONCLUSION

The new systems used for the redesign of MRCTS are now all functional. Unfortunately, a few problems occurred during the implementation, which slowed the entire processing. However, once everything is resolved, this redesign will be able to reach its objectives which are to produce better and more timely results at a lower cost.

Work still has to be completed before the redesign is finalized. As mentioned previously, the hook-up of MRCTS with the BR must still be completed; this will improve the BR information that is used to update the MRCTS frame. The redesign, implemented with the production of the May 1995 estimates, caused a break between the old series (pre-May 1995) and the new series. A link has to be establish between the two series to be able to make relevant year-to-year comparison and to eliminate the impact of the redesign in the comparison.

Finally, all the system components of the survey (sample selection, edit, imputation and estimation) developed by the methodologists are presently ran by the methodologists servicing MRCTS. In the near future, all this work will be transferred to the subject-matter specialists. They will then be able to handle all the aspects of their survey, and the methodologists will be available to assist them with the review and analysis of their survey as well as look at potential improvements to the survey process.

ACKNOWLEDGEMENTS

The authors wish to thank the reviewers for their helpful comments and suggestions.

REFERENCES

Hidiroglou, M.A., Bennett, R., Eady, J. and Maisonneuve, L. (1980). Sample design of the Monthly Restaurants, Caterers and Taverns Survey. Survey Methodology, No. 1, 57-83.

Business Register Division Newsletter, Volume 1, No. 1.

Internal Statistics Canada papers on the Generalized Systems.